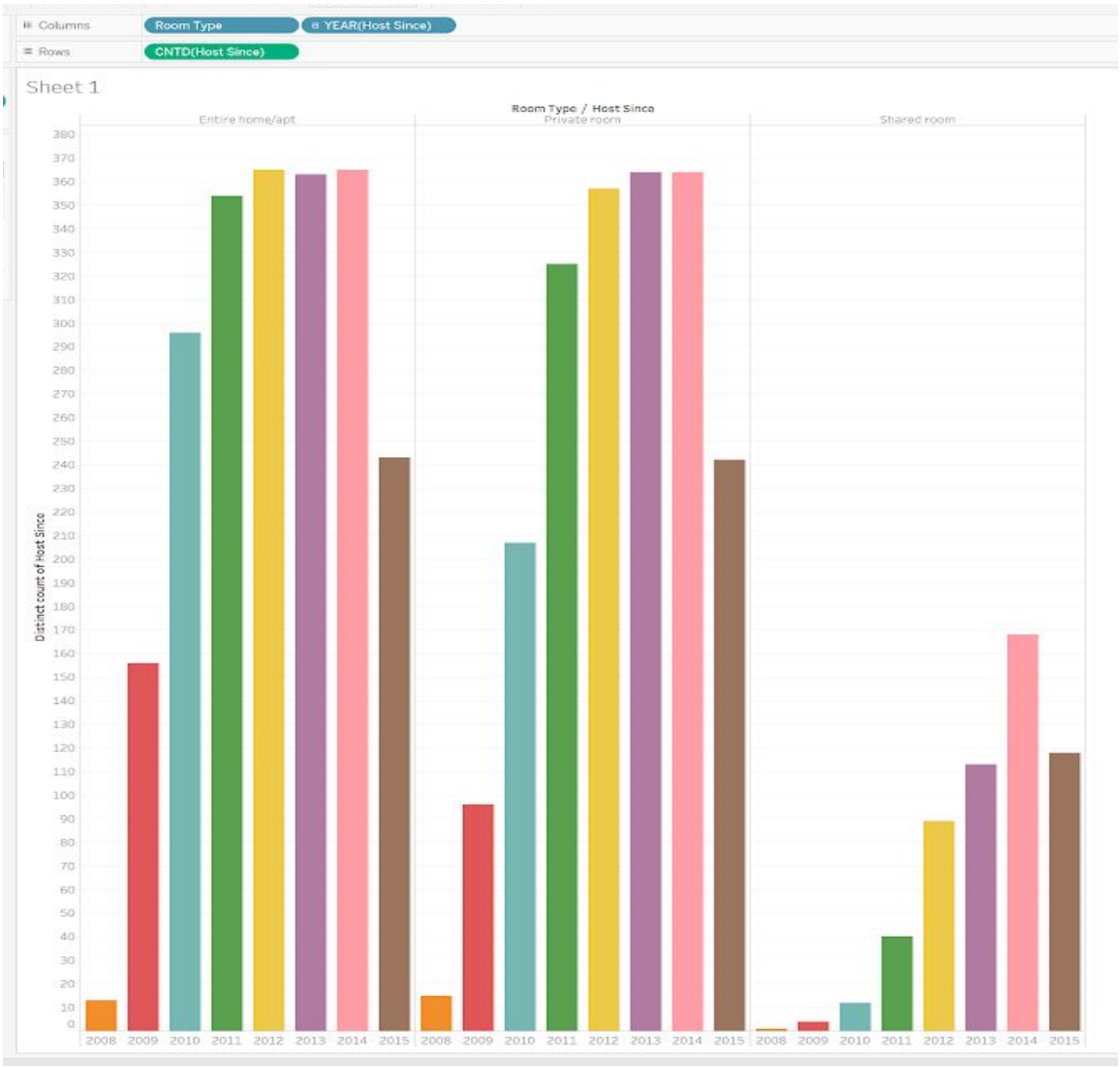
Homework 4: Use and Critique Tableau - Airbnb Dataset

By Kevin Lieu and Pranshav Thakkar

Part 1: Report

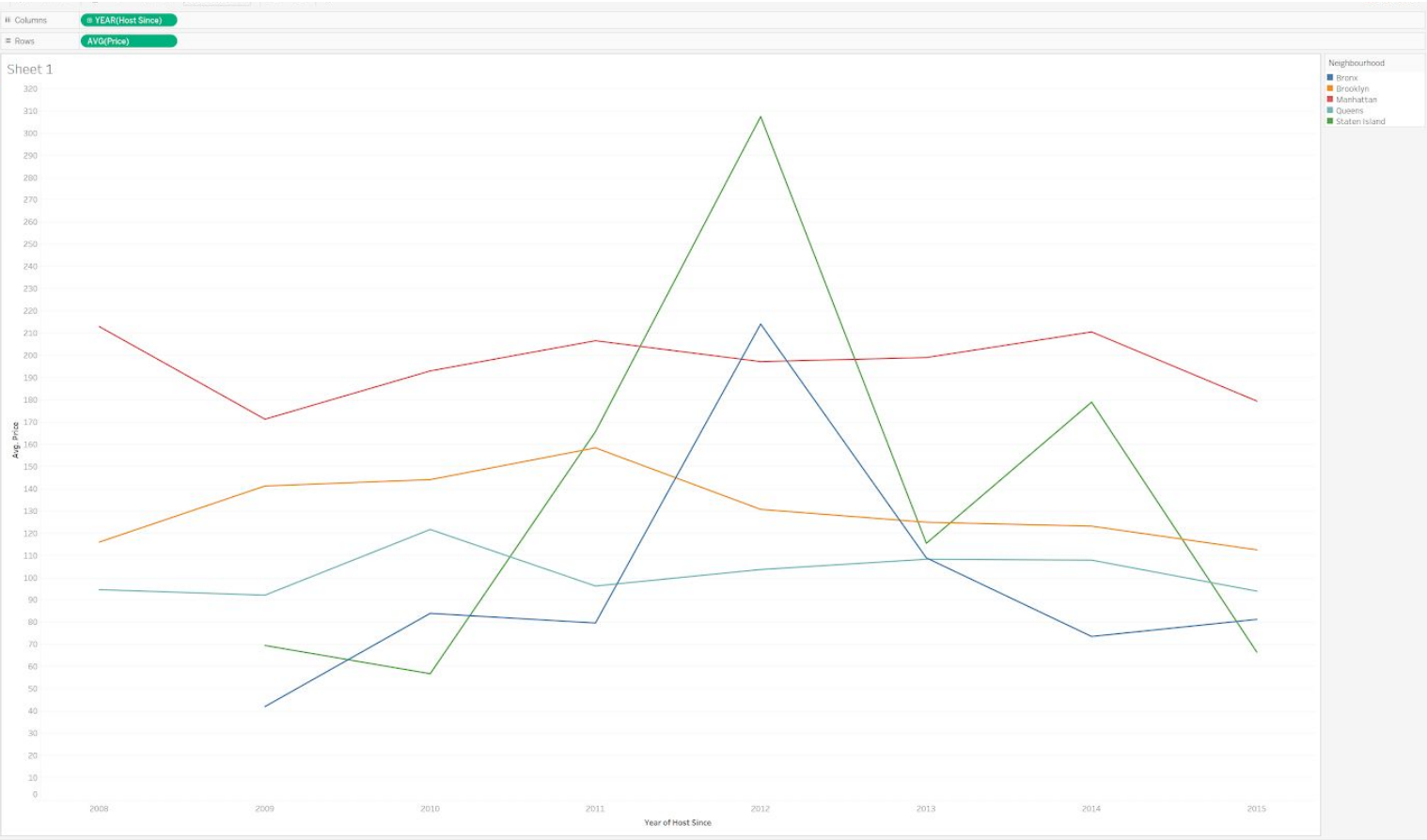
Question 1: From 2008 to 2015, how many new people have joined airbnb as a host across different room types?



Our visualization is separated by the three different room types and shows how many people have joined Airbnb each year. By comparing the length of each bar across all room types, we can see that there is a similar trend across all room types. It appears that Airbnb wasn't very popular

back in 2008 but started gaining popularity in 2009. Following 2009, there was a steady increase in host interest with interest finally slowing down in 2015. Perhaps with the constant increase before 2015, the supply of hosts has finally reached an equilibrium and settled down.

Questions 2: For each neighborhood, how has the average price of an airbnb changed over the years?



Based on the average price of Airbnbs and the year they started hosting, we can see the difference between neighborhoods in New York and how they compare throughout the years. Overall, we can tell that Manhattan, Brooklyn, and Queens have maintained a steady price rate through the years since people have been hosting. However, with the Bronx and Staten Island, there were unexpected price increases from 2010 to 2012. Especially considering Staten Island, there were significant price spikes in 2012 and 2014 that seem incredibly unusually considering

the prices fall down the following years. The visualization really brings into question what could have potentially made property in Staten Island and the Bronx increase so much within the span of these years. Additionally, this lets us easily tell the relative cost of each neighborhood with respect to the other available renting locations.

Question 3: Does a higher price mean a higher review score rating or vice-versa? Additionally, does a higher number of reviews mean a higher review score?

In this visualization, we try to compare the prices of an Airbnb to their review score rating and then those score ratings to the number of reviews given. Each individual dot is representative of a single Airbnb property that has hosted and been rated before. As such, in terms of pricing, there does not seem to be any relative relationship between how much a property cost to rent and their rating. There are plenty of high ratings given throughout both the lower-priced properties and the average-priced properties compared to the more expensive ones. However, there does appear to be an association between the relative rating score and the quantity of reviews a property

receives. As the review rating score increases, we can see that the individual dots get bigger and appear more frequently. This implies that the higher the rating a property has the more likely they are to receive additionally reviews to the ones they already have.

Part 2: Critique

Tableau is a great system for user tasks that include comparing and contrasting given numerical data, finding answers to specific questions about numerical data sets, and exploring a dataset visually. It is good at allowing users to simply drag and drop different dimensions that it automatically generates from the data and set them up as either an axes or a data mark of a visualization. Tableau has an easy-to-use selection for different types of visualizations that conveniently show how many dimensions and measures it will need to generate that type of viz, allowing users to quickly preview different kinds of visualizations to see which one displays their data best. The visualization techniques are adaptive and reflect changes automatically, which is a bonus when users are exploring their data to find some sort of correlation between different dimensions and measures. Another useful function is the ability to change dimensions to measure by calculating the average or sum of the dimension, or changing a measure into a dimension. This allows the user to create visualizations that answer questions based on the data itself rather than just based on the data labels.

A drawback about Tableau is that it does not offer an easy-to-use functionality to parse data entries with respect to their ordinal, nominal, or quantitative values. It was a struggle to identify which attributes represented what values and how to manipulate them more intuitively. For example, if we were looking for data in the form of “Team A beat Team B via penalties”, in a soccer dataset, then Tableau would simply show data directly from the spreadsheet without allowing more ways of manipulating that data such as stacking attribute filters through multi-selection. In addition, when switching between the various visualization techniques, the system often changes or drops some dimensions/measures, which can make it hard for users to keep track of what is being displayed to them.

Overall, Tableau is an excellent program for providing visualizations such as bar charts and line graphs to compare and contrast given fields of a dataset, with a robust sorting and filtering system for customization. These can easily be used to analyze and discover new information and relationships within any dataset. The system also has a couple of different scatter-plot options that are helpful in analyzing aggregate forms of data and comparing multiple fields. Tableau smartly adapts these visualizations to accommodate scenarios in which the users want to present a multivariate information with a variety of visualizations that are easily adaptable using the drag and drop features. It creates a code-less environment for experimentation without having to worry about graph reconstruction on every instance of change. Thus, while there may be some UI annoyances and certain difficulties in data manipulation, Tableau makes for a great tool to easily create and change complex visualizations.